CS145: Probability & Computing

Lecture 13: Covariance, Bivariate and Multivariate Normal Distributions

Instructor: Eli Upfal

Brown University Computer Science

Figure credits:
Bertsekas & Tsitsiklis, *Introduction to Probability*, 2008
Pitman, *Probability*, 1999
CS145: Lecture 12-13 Outline

- Covariance and Correlation
- Bivariate Normal Distributions
- Multivariate Normal Distributions
  Advanced topic not covered in homeworks or exams!
  Requires linear algebra for in-depth study.
Variance of Sums of Random Variables

- If $Z = X + Y$ for (possibly dependent) random variables $X$ and $Y$:
  \[ E[Z] = E[X] + E[Y] \]

- To simplify analysis, define “centered” random variables:
  \[ \tilde{X} = X - E[X], \quad \tilde{Y} = Y - E[Y] \]

- The variance of $Z$ is then equal to:
  \[ \text{Var}[Z] = E[(Z - E[Z])^2] = E[(\tilde{X} + \tilde{Y})^2] \]
  \[ \text{Var}[Z] = \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}[X, Y] \]

- The covariance of $X$ and $Y$ is defined as:
By definition, the covariance of random variables $X$ and $Y$ equals:


Intuition via “centered” random variables:

$$\tilde{X} = X - E[X], \quad E[\tilde{X}] = 0.$$  
$$\tilde{Y} = Y - E[Y], \quad E[\tilde{Y}] = 0.$$  

$$\text{Cov}(X, Y) = E[\tilde{X}\tilde{Y}]$$

Independent variables have zero covariance:

$$\text{Cov}(X, Y) = E[\tilde{X}\tilde{Y}] = E[\tilde{X}]E[\tilde{Y}] = 0$$

if $f_{XY}(x, y) = f_X(x)f_Y(y)$
The covariance depends on units of variables X and Y

Often convenient to use **standardized variables**:

\[ \tilde{X} = \frac{X - \mu_x}{\sigma_x} \quad \tilde{Y} = \frac{Y - \mu_y}{\sigma_y} \]

\[ \mu_x = E[X], \mu_y = E[Y] \]
\[ \sigma_x^2 = \text{Var}(X), \sigma_y^2 = \text{Var}(Y) \]

For these standardized variables, we have changed the "coordinate system" or "units" so that:

\[ E[\tilde{X}] = 0, \quad \text{Var}(\tilde{X}) = 1 \]
\[ E[\tilde{Y}] = 0, \quad \text{Var}(\tilde{Y}) = 1 \]
Correlation Coefficient

- The covariance depends on units of variables $X$ and $Y$
- Often convenient to use **standardized variables**:
  \[
  \tilde{X} = \frac{X - \mu_x}{\sigma_x} \quad \tilde{Y} = \frac{Y - \mu_y}{\sigma_y}
  \]
  \[
  \mu_x = E[X], \mu_y = E[Y] \quad \sigma_x^2 = \text{Var}(X), \sigma_y^2 = \text{Var}(Y)
  \]

- The **correlation coefficient** “rho” is defined to equal:
  \[
  \rho(X, Y) = E[\tilde{X}\tilde{Y}] = E \left[ \left( \frac{X - \mu_x}{\sigma_x} \right) \cdot \left( \frac{Y - \mu_y}{\sigma_y} \right) \right] = \frac{\text{Cov}(X, Y)}{\sigma_x \sigma_y}
  \]

- For any joint distribution, we have $-1 \leq \rho(X, Y) \leq 1$

Proof: Cauchy-Swartz Inequality
Empirical Correlation Coefficients

**Correlation coefficient of empirical distribution** of $N$ observations:

$$
\rho = \frac{1}{N} \sum_{i=1}^{N} \frac{(x_i - \mu_x)(y_i - \mu_y)}{\sigma_x \sigma_y}
$$

$$
\mu_x = \frac{1}{N} \sum_{i=1}^{N} x_i
$$

$$
\sigma_x^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu_x)^2
$$
Empirical Correlation Coefficients

Correlation coefficient of empirical distribution of $N$ observations:

<table>
<thead>
<tr>
<th></th>
<th>1.0</th>
<th>0.8</th>
<th>0.4</th>
<th>0.0</th>
<th>-0.4</th>
<th>-0.8</th>
<th>-1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Dependence grows stronger as $\rho$ approaches $-1$ or $+1$:

  If $\rho = +1$ then $(X - \mu_x) = c(Y - \mu_y)$ for some $c > 0$.
  If $\rho = -1$ then $(X - \mu_x) = c(Y - \mu_y)$ for some $c < 0$. 
Empirical Correlation Coefficients

Correlation coefficient of empirical distribution of \( N \) observations:

\[
\begin{array}{cccccccc}
1.0 & 0.8 & 0.4 & 0.0 & -0.4 & -0.8 & -1.0 \\
\end{array}
\]

Example empirical statistics of real data:

- Fathers: mean height: 5′9″, SD: 2″
- Sons: mean height: 5′10″, SD: 2″
- correlation: 0.5

Karl Pearson’s study of 1078 father, son pairs (~1900).
Data from Pitman Sec. 6.5.
Empirical Correlation Coefficients

Correlation coefficient of empirical distribution of $N$ observations:

Example empirical statistics of real data:

<table>
<thead>
<tr>
<th>Correlation Coefficient</th>
<th>Verbal</th>
<th>Math</th>
<th>Writing</th>
<th>Verbal + Math</th>
<th>Verbal + Math + Writing</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSAT Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal</td>
<td>0.817</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math</td>
<td></td>
<td>0.816</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Writing</td>
<td></td>
<td></td>
<td>0.761</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal + Math</td>
<td></td>
<td></td>
<td></td>
<td>0.858</td>
<td></td>
</tr>
<tr>
<td>Verbal + Math + Writing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.866</td>
</tr>
</tbody>
</table>

Empirical Correlation Coefficients

Correlation coefficient of empirical distribution of $N$ observations:

1.0  0.8  0.4  0.0  −0.4  −0.8  −1.0

1.0  1.0  1.0  0.0  −1.0  −1.0  −1.0

0.0  0.0  0.0  0.0  0.0  0.0  0.0

WARNING: Zero correlation does not imply independence!!!
Covariance and Correlation

Bivariate Normal Distributions

Multivariate Normal Distributions

Advanced topic not covered in homeworks or exams!
Requires linear algebra for in-depth study.
Normal Random Variables

\[ f_X(x) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2} \]

\[ E[X] = \mu \]

\[ \text{Var}[X] = E[(X - \mu)^2] = \sigma^2 \]

\[ \sqrt{\text{Var}[X]} = \sigma \text{ is the standard deviation} \]

**Theorem**: A linear function of a Gaussian variable is Gaussian!

\[ Y = aX + b \]

\[ f_Y(y) = \frac{1}{\sqrt{2\pi}\bar{\sigma}^2} e^{-\frac{1}{2} \left( \frac{y-\bar{\mu}}{\bar{\sigma}} \right)^2} \]

\[ f_Y(y) = \frac{1}{|a|} f_X \left( \frac{y-b}{a} \right) \]

\[ \bar{\mu} = a\mu + b, \quad \bar{\sigma} = |a|\sigma \]
Standard Normal Random Variables

- If $X \sim N(\mu, \sigma^2)$ then for any constants $a$ and $b$ the random variable $aX + b$ is distributed $N(a\mu + b, a^2\sigma^2)$.
- If $X \sim N(\mu, \sigma^2)$ then $Z = \frac{X - \mu}{\sigma}$ is distributed $N(0, 1)$.
- $N(0, 1)$ is the standard Normal distribution.

$$Pr(Z \leq z) = \Phi_Z(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt$$

$$\phi_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$
Two Independent Normal Variables

The set of points where \( f_{XY}(x,y) = c \), for any constant \( c \), is a circle centered at the mean.
Two Independent Normal Variables

The set of points where \( f_{X,Y}(x,y) = c \), for any constant \( c \), is an ellipse centered at the mean.

\[
\text{Var}[X] = \sigma_x^2 = \sigma_y^2 = \text{Var}[Y],
\]

\[
\text{Var}[X] = \sigma_x^2 < \sigma_y^2 = \text{Var}[Y].
\]
Bivariate Normal Distribution

\[ f_U(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \quad f_V(v) = \frac{1}{\sqrt{2\pi}} e^{-v^2/2} \]

A bivariate normal distribution is any joint distribution defined as a linear function of two independent normal distributions

First consider the following particular linear function:

\[ X = \sqrt{\frac{1 + \rho}{2}} U + \sqrt{\frac{1 - \rho}{2}} V \]
\[ Y = \sqrt{\frac{1 + \rho}{2}} U - \sqrt{\frac{1 - \rho}{2}} V \]

The variables \( X \) and \( Y \) are Gaussian with statistics:

\[ E[X] = E[Y] = 0 \quad \text{Var}(X) = \text{Var}(Y) = 1 \]
\[ \rho(X, Y) = \text{Cov}(X, Y) = \rho \]
Bivariate Normal Density Functions

$\rho = 1.0 \quad 0.8 \quad 0.4 \quad 0.0 \quad -0.4 \quad -0.8 \quad -1.0$

$\rho > 0 \quad \rho = 0 \quad \rho < 0$
A **bivariate normal distribution** is any joint distribution defined as a *linear function of two independent normal distributions*

First consider the following particular linear function:

\[
X = \sqrt{\frac{1 + \rho}{2}} U + \sqrt{\frac{1 - \rho}{2}} V \\
Y = \sqrt{\frac{1 + \rho}{2}} U - \sqrt{\frac{1 - \rho}{2}} V
\]

The **joint probability density function** of \( X \) and \( Y \) equals:

\[
f_{XY}(x, y) = \frac{1}{2\pi \sqrt{1 - \rho^2}} \exp \left\{ -\frac{x^2}{2(1 - \rho^2)} - \frac{y^2}{2(1 - \rho^2)} + \frac{\rho xy}{1 - \rho^2} \right\}
\]

\( \rho = 0 \implies f_{XY}(x, y) = \frac{1}{2\pi} \exp \left\{ -\frac{x^2}{2} - \frac{y^2}{2} \right\} = f_X(x)f_Y(y) \implies \text{Independence!} \)
Bivariate Normal Distribution

Consider two independent “standard” normal variables

\[ f_U(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \quad f_V(v) = \frac{1}{\sqrt{2\pi}} e^{-v^2/2} \quad E[U] = E[V] = 0 \]
\[ \text{Var}(U) = \text{Var}(V) = 1 \]

We construct two new Gaussian variables via a linear function:

\[ X = aU + bV + c \]
\[ Y = dU + eV + f \]

The joint probability density of \( X, Y \) is then bivariate normal:

\[ f_{XY}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp \left\{ -\frac{(x - \mu_x)^2}{2\sigma_x^2(1-\rho^2)} - \frac{(y - \mu_y)^2}{2\sigma_y^2(1-\rho^2)} + \frac{\rho(x - \mu_x)(y - \mu_y)}{\sigma_x\sigma_y(1-\rho^2)} \right\} \]

\[ \mu_x = E[X], \mu_y = E[Y] \quad \sigma_x^2 = \text{Var}(X), \sigma_y^2 = \text{Var}(Y) \quad \rho = \frac{\text{Cov}(X, Y)}{\sigma_x\sigma_y} \]
Interpreting Normal Parameters

\[ f_{XY}(x, y) = \frac{1}{2\pi \sigma_x \sigma_y \sqrt{1-\rho^2}} \exp \left\{ -\frac{(x-\mu_x)^2}{2\sigma_x^2(1-\rho^2)} - \frac{(y-\mu_y)^2}{2\sigma_y^2(1-\rho^2)} + \frac{\rho(x-\mu_x)(y-\mu_y)}{\sigma_x \sigma_y (1-\rho^2)} \right\} \]

- Coordinate system and units for random variable \( X \):
  - Mean: \( \mu_x = E[X] \)
  - Standard deviation: \( \sigma_x = \sqrt{\text{Var}(X)} \)
  - \( P(X \leq \mu_x) = P(X \geq \mu_x) = 0.5 \)

- Coordinate system and units for random variable \( Y \):
  - Mean: \( \mu_y = E[Y] \)
  - Standard deviation: \( \sigma_y = \sqrt{\text{Var}(Y)} \)
  - \( P(Y \leq \mu_y) = P(Y \geq \mu_y) = 0.5 \)

- Dependence between \( X, Y \) measured by correlation coefficient:
  \[ \rho = \frac{\text{Cov}(X, Y)}{\sigma_x \sigma_y}, \quad -1 \leq \rho \leq 1 \]

For normal variables: \( X \) and \( Y \) independent if and only if \( \rho = 0 \)
Two Correlated Normal Variables

\[
f_{XY}(x, y) = \frac{1}{2\pi \sigma_x \sigma_y \sqrt{1 - \rho^2}} \exp \left\{ -\frac{(x - \mu_x)^2}{2\sigma_x^2(1 - \rho^2)} - \frac{(y - \mu_y)^2}{2\sigma_y^2(1 - \rho^2)} + \frac{\rho(x - \mu_x)(y - \mu_y)}{\sigma_x \sigma_y(1 - \rho^2)} \right\}
\]
Covariance and Correlation

Bivariate Normal Distributions

Multivariate Normal Distributions

Advanced topic not covered in homeworks or exams!
Requires linear algebra for in-depth study.
Multivariate Normal Distribution

Let $X^T = (X_1, \ldots, X_n)$ be a vector of $n$ independent, standard normal random variables. $E[X_i] = 0$ and $Var[X_i] = 1$.

Let $Y^T = (Y_1, \ldots, Y_m)$ be random variable vector obtained by a linear transformation on the vector $X^T$:

\[
Y_1 = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1n}X_n + \mu_1;
\]

\[
Y_2 = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2n}X_n + \mu_2;
\]

\[
\vdots
\]

\[
Y_m = a_{m1}X_1 + a_{m2}X_2 + \cdots + a_{mn}X_n + \mu_m.
\]

Let $A$ denote the matrix of coefficients $a_{ij}$, and $\bar{\mu}^T = (\mu_1, \ldots, \mu_m)$. Then we can write

\[
Y = AX + \bar{\mu}.
\]
Mean Vectors & Covariance Matrices

\[ Y_1 = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1n}X_n + \mu_1; \]
\[ Y_2 = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2n}X_n + \mu_2; \]
\[
\vdots
\]
\[ Y_m = a_{m1}X_1 + a_{m2}X_2 + \cdots + a_{mn}X_n + \mu_m. \]

\[ Y = AX + \bar{\mu}, \quad E[Y_i] = \mu_i, \quad Var[Y_i] = \sum_{j=1}^{n} a_{i,j}^2, \quad E[\bar{Y}] = \bar{\mu}, \]

\[ Cov(Y_i, Y_j) = \sum_{k=1}^{n} a_{i,k}a_{j,k}. \]

The covariance matrix for \( Y \) is given by

\[ \Sigma = \mathbf{A} \mathbf{A}^T = \begin{pmatrix} Var[Y_1] & Cov(Y_1, Y_1) & \cdots & Cov(Y_1, Y_n) \\ Cov(Y_1, Y_2) & Var[Y_2] & \cdots & Cov(Y_2, Y_n) \\ \vdots & \vdots & \ddots & \vdots \\ (Y_m, Y_1) & Cov(Y_m, Y_2) & \cdots & Var[Y_m] \end{pmatrix} = E[(Y - \bar{\mu})(Y - \bar{\mu})^T]. \]
Joint Multivariate Normal Distribution

If $A$ has a full rank, then $X = A^{-1}(Y - \bar{\mu})$, and we can derive a density function for the joint distribution.

$$\Pr(Y \leq y) = \Pr(Y - \mu \leq y - \mu)$$
$$= \Pr(AX \leq y - \mu)$$
$$= \Pr(X \leq A^{-1}(y - \mu))$$
$$= \frac{1}{(2\pi)^{n/2}} \int_{\bar{\omega} \leq A^{-1}(y-\mu)} e^{-\frac{\bar{\omega}^T \bar{\omega}}{2}} d\omega_1 \ldots d\omega_n.$$  

Changing the integration variables to $\bar{z} = A\bar{\omega} + \bar{\mu}$ we have

$$Pr(Y \leq y) = \frac{1}{\sqrt{(2\pi)^n |AA^T|}} \int_{-\infty}^{y_1} \ldots \int_{-\infty}^{y_n} e^{-\frac{1}{2}(\bar{z} - \mu)^T (A^{-1})^T A^{-1} (\bar{z} - \mu)} d\bar{z}_1 \ldots d\bar{z}_n.$$  

Here $|AA^T|$ denotes the determinant of $AA^T$, a term which arises under the multivariate change of variables.
Applying \((A^{-1})^T A^{-1} = (A^T)^{-1} A^{-1} = (AA^T)^{-1} = \Sigma^{-1}\), we can write the distribution function of \(Y\) as

\[
\Pr(Y \leq \bar{y}) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} \int_{-\infty}^{y_1} \cdots \int_{-\infty}^{y_n} e^{-\frac{1}{2}(\bar{z} - \mu)^T \Sigma^{-1} (\bar{z} - \mu)} \, dz_1 \cdots \, dz_n \quad (1)
\]

where, again,

\[
\Sigma = AA^T = E[(Y - \mu)(Y - \mu)^T].
\]
A vector $Y^T = (Y_1, \ldots, Y_n)$ has a multivariate normal distribution, denoted $Y \sim N(\bar{\mu}, \Sigma)$, if and only if there is an $n \times k$ matrix $A$, a vector $X^T = (X_1, \ldots, X_k)$ of $k$ independent standard normal random variables, and a vector $\bar{\mu}^T = (\mu_1, \ldots, \mu_n)$, such that

$$Y = AX + \bar{\mu}.$$ 

If $\Sigma = AA^T = E[(Y - \bar{\mu})(Y - \bar{\mu})^T]$ has full rank, then the density of $Y$ is

$$\frac{1}{\sqrt{(2\pi)^n|\Sigma|}} e^{-\frac{1}{2}(Y - \bar{\mu})^T \Sigma^{-1} (Y - \bar{\mu})}.$$ 

If $\Sigma$ is not invertible then the joint distribution has no density function.
Multivariate Normal Probability Density

\[
\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp \left\{ -\frac{1}{2\sigma^2}(x - \mu)^2 \right\}
\]

\[
\mathcal{N}(x|\mu,\Sigma) = \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu) \right\}
\]

\[
\mu = E[X]
\]

\[
\Sigma = E[(X - \mu)(X - \mu)^T]
\]

D-dimensional ellipsoids parameterized by mean vector & covariance matrix